

PTE Közgazdaságtudományi Kar és POTE Egészségügyi Szerve-  
zési Intézet

Cardiovascularis kérdőíves előszűrés adatainak számítógépes elem-  
zése, standardizált metodika kidolgozása

Csébfalvi Gy. és Szilárd I.P.

Az egészségügyünket jellemző preventív szemlélet többek között a lakosság egészére kiterjedő, rendszeres, komplex szűrővizsgálatok bevezetését is jelenti. A jelenlegi adottságokat - objektív és szubjektív feltételeket - tekintve ez klasszikus módszerekkel: tényleges orvosi vizsgálattal ma még megvalósíthatatlan.

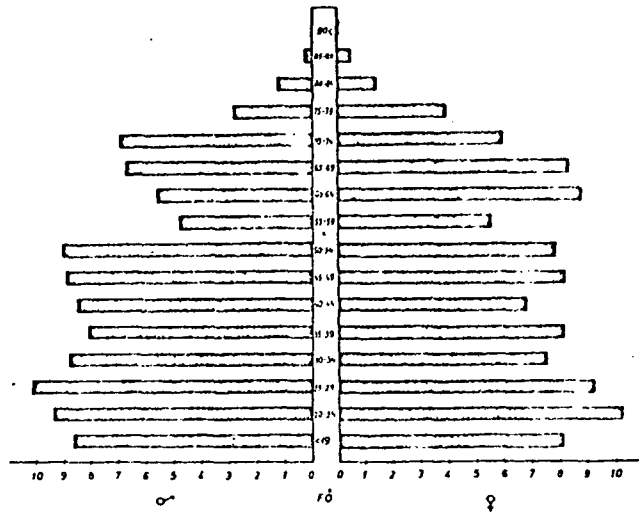
A POTE Egészségügyi Szervezési Intézetében évek óta folytatunk olyan céltudatos kutatómunkát, mely a szűrővizsgálatokkal szemben támasztott alapvető követelmények megtartása mellett (nagy szenzitivitás, megfelelő specificitás), a magasan kvalifikált orvosi munkaóra-igényt csökkenti, illetőleg ennek effektivitását növeli. A kérdőíves előszűrés (6) Intézetünkben kialakított módszere lehetővé teszi egészségügyi középkezelőknél nagyobb mérvű bevonását a vizsgálatba, s ezen túlmenően az értékelés és a sürgősségi sorrend felállítása számítógépes adatfeldolgozással történhet.

A cardiovascularis betegségcsoport, valamint a kapcsolódó pulmo- és nephropathiák magas részesedési aránya az összmorbiditásból indokoltá teszi a speciálisan csak erre a betegségcsoportra koncentrált előszűrés kísérleti jellegű elvégzését.

Az előszűrést a Pécs-Rácvárosi (44.sz.) orvosi körzetben végeztük. Az előkészítés a szűrés lebonyolításának megszervezésén kívül az érintett lakosság értesítését, valamint a szűrés céljáról való felvilágosítását is magában foglalta.

A vizsgált lakosság kor és nem szerinti megoszlását az 1. ábrán szemléltetjük. A vizsgálatban résztvett 1116 nő és 965 férfi. A

PÉCS RÁCVÁROSI ORVOSI KÖRZET FELNŐTT LAKOSSÁGÁNAK  
KOR ÉS NEM SZERINTI MEGOSZLÁSA (1975)



1. ábra

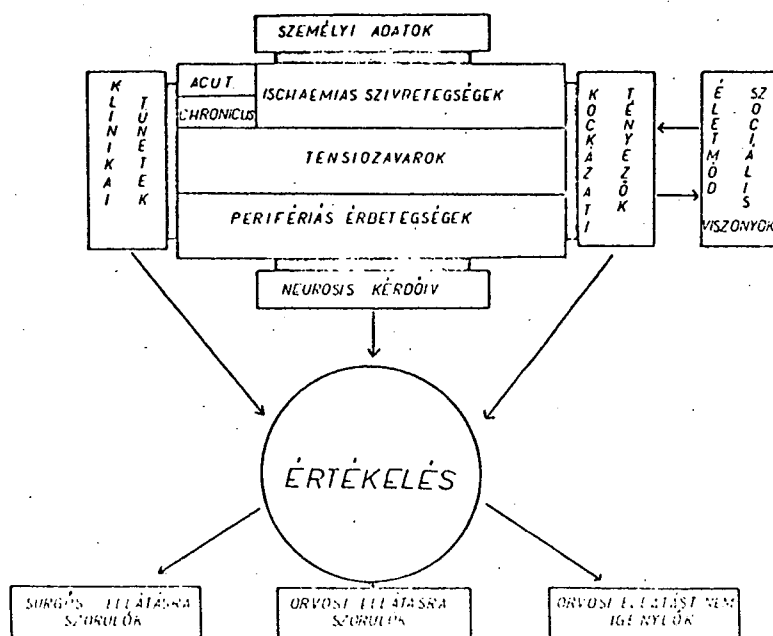
nők átlagos életkora 45.1 év, míg a férfiaké 43,7 év volt. A 2081 szűrővizsgálati lapból további feldolgozásra 2070 volt alkalmas.

Magát az előszűrést V. évf. orvostanhallgatók végezték. Előzetes értesítés után a vizsgáltkákat lakásukon keresték fel, szükség esetén segédkeztek a kérdőív megválaszolásánál, s egyben vérnyomásmérést is végeztek. Ennek eredményét a kérdőíven rögzítették. A vérnyomást 5 perc ülőhelyzetben töltött nyugalmi periódus után, az anaeroid vérnyomásmérővel mérték.

A szűrésre kidolgozott komplex kérdőív felépítését a 2. ábra szemlélteti.

A betegségcsoportba tartozó kórképek klinikai tüneteinek kívül rákérdeztünk a feltételezett kockázati tényezőkre is, a párhuzamosan végzett psychodiagnosztikai tesztként kitöltöttük az Os-váth-féle 20 pontos önjellemző panaszleltárt. Az értékelés első megközelítésben csak a szubjektív klinikai panaszokra adott válaszok alapján történt, melyeket becslés, illetőleg klinikai tapasztalat alapján egy-, két- és háromjegyű számmal súlyoztunk.

# KÉRDŐÍVES ELŐSZÜRÉS ÉRTEKELÉSI VÁZLATA



2. ábra

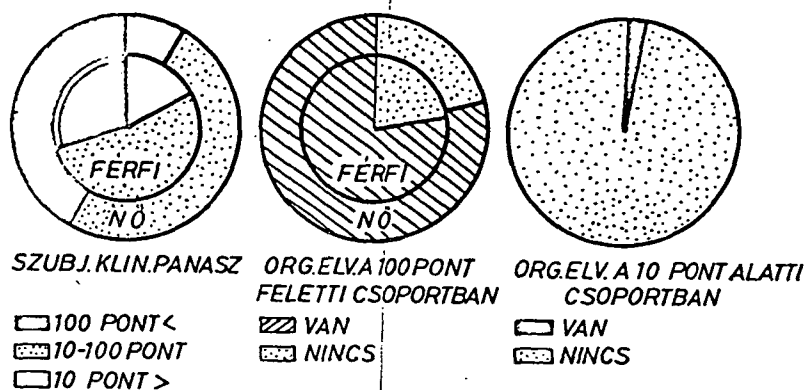
Munkahipotézisünk szerint az összérték alapján háromjegyű kategóriába jutó esetek rövid időn belül orvosi vizsgálatot igényelnek.

Az első - még nem gépi adatfeldolgozással végzett - értékelés alapján sürgős orvosi vizsgálatot igénylők 74 %-át, s az egyértelműen egészségesnek nyilvánítottak 20 %-át felülvizsgáltuk (3. ábra), s ennek alapján a kérdőív validitását, azaz a kapott korrekt értékeknek az összvizsgálaton belüli százalékos arányát 88 %-osnak, szenzitivitását 96 %-osnak, specificitását 82 %-osnak találtuk (4. ábra).

A kérdőíves előszűrési értékelése matematikai modelljének kidolgozásánál az orvosilag felülvizsgált 773 fős populáció adataira támaszkodtunk. A továbbiakban ezt ismertetjük.

A matematikai modell diszkriminancia analízisen (DA) alapul. A DA két populáció szétválasztására alkalmas eljárás több kvantitatív, illetőleg kvalitatív tulajdonság együttes figyelembevételével. A DA programjai FORTRAN nyelven készültek, az ESZ/DOS operációs rendszer felügyelete alatt R-20-as berendezésen futottak.

**A SZUBJEKTIV KLINIKAI PANASZOK SÚLYOSSÁGA,  
VALAMINT AZ OBJEKTIV KLINIKAI ELVÁLTOZÁS %-OS  
GYAKORISÁGA A VIZSGÁLT LAKOSSÁG KÖRÉBEN**



3. ábra

**A CARDIOVASCULARIS ELŐSZÜRÉS KÉRDŐÍVÉNEK  
MEGBIZHATÓSÁGI VIZSGÁLATA (100 FŐRE SZÁMITVA)**

ORVOSI VIZSG. TESZT	POZITIV	NEGATIV	$\Sigma$
POZITIV	79	21	100
NEGATIV	3	97	100
$\Sigma$	82	118	200

VALIDITÁS: 88 %  
SENZIBILITÁS: 96 %  
SPECIFICITÁS: 82 %

4. ábra

A  $P_1$  (a felülvizsgálat szerint egészséges) és a  $P_2$  (a felülvizsgálat szerint beteg) populáció elemszáma 305, illetve 468 fő.

A módszertani vizsgálat a szubjektív klinikai panaszokra és a gyógyszeresedésre irányuló kérdésekre adott válaszok, valamint a vérnyomásmérés eredményéből adódó, összesen 28 adat figyelembevételével történt. Az egyes adatokat binárisan kódoltuk, illetőleg binarizáltuk. A vizsgálat során az orvosilag felülvizsgált lakosokat az

$$\underline{X} = (X_1, X_2, \dots, X_m), y$$

adatokkal jellemeztük, ahol

$\underline{X}$  a binárisan kódolt válaszok vektora (a továbbiakban válaszvektor), és

$y$  a felülvizsgálat eredményét jelző kvalitatív változó.

$y = \begin{matrix} 1 & \text{ha a felülvizsgálat eredménye negatív} \\ 2 & \text{ha a felülvizsgálat eredménye pozitív.} \end{matrix}$

DA segítségével a következő kérdésre kerestünk választ:

1. A válaszvektorok alapján kimutatható-e szignifikáns különbség a két populáció között?

2. Milyen tévedési valószínűséggel lehet a válaszvektorokból a felülvizsgálat eredményére következtetni?

3. Az egyes kérdésekre adott válaszok - egymáshoz viszonyítva - mennyire jelentősek a két populáció szétválasztása szempontjából?

4. Milyen eljárás segítségével lehet a legkisebb tévedési valószínűséggel eldönteni, hogy egy újabb megfigyelt személy az egészséges, illetőleg a beteg populációba sorolandó-e?

A számítógépes elemzés eddigi szakaszában vizsgálataink alapvetően a felvétel során szerepeltetett kérdések diszkrimináló jellegének meghatározására, és ennek alapján új, nagyobb validitású kérdőív kidolgozására irányultak.

A vizsgálat során öt eljárás alkalmazhatóságát vizsgáltuk. Az eljárásokat az jellemzi, hogy viszonylag nagy változószámig számítástechnikailag jól kezelhetők, és nem igényelnek iteratív becslési technikát.

A vizsgált módszerek közül kettő az un. Bahadur-féle reparametrizáción alapul, tehát kifejezetten bináris változók vizsgálatára alkalmas. Jelölje  $\pi_i(\underline{X})$ , ( $i=1,2$ ) annak a valószínűségét, hogy az  $\underline{X} = (X_1, X_2, \dots, X_m)$  válaszvektorral jellemzett személy az  $i$ -edik ( $i=1,2$ ) populációba tartozik. Bahadur (1) megmutatta, hogy a  $\pi_i(\underline{X})$ , ( $i=1,2$ ) multinomiális valószínűség a következőképpen állítható elő:

Legyen:  $\underline{X} = (X_1, X_2, \dots, X_m)$  a válaszvektornak megfelelő valószínűségi vektorváltozó,  $p_{ij} = E_i(X_j)$  az  $X_j$  változó várható értéke az  $i$ -edik populációban ( $i=1,2$ ,  $j=1,2,\dots,m$ ),

$$Z_{ij} = (X_j - p_{ij}) / \sqrt{p_{ij}(1-p_{ij})},$$

$$r_{ij}(jk) = E(Z_{ij} Z_{ik}),$$

$$r_{ij}(jkl) = E(Z_{ij} Z_{ik} Z_{il}),$$

$$\vdots$$

$$r_{ij}(12\dots m) = E(Z_{i1} Z_{i2} \dots Z_{im}).$$

Ekkor

$$\pi_i(\underline{X}) = \prod_{j=1}^m \frac{X_j^{p_{ij}} (1-X_j)^{1-p_{ij}}}{p_{ij}^{p_{ij}} (1-p_{ij})^{1-p_{ij}}} \left[ 1 + \sum_{j < k} r_{ij}(jk) Z_{ij} Z_{ik} + \dots \right],$$

A Bahadur-féle reparametrizáció a  $\pi_i(\underline{X})$ , ( $i=1,2$ ) multinomiális valószínűséget

- a  $p_{ij}$  várható értékek, illetőleg
- a változók közötti sztohasztikus kapcsolatokat jellemző  $r_{ij}(jk)$  korrelációs együtthatók függvényeként adja meg.

A Bahadur-féle reparametrizáción alapuló eljárások a  $\pi_i(\underline{X})$ , ( $i=1,1$ ) multinomiális valószínűségeket első, illetőleg másodrendű formulával közelítik. (Az eljárásokra a továbbiakban a B1, illetve a B2 jelölésekkel hivatkozunk.)

Elsőrendű közelítő formulát kapunk, ha az összefüggésben a korrelációs együttthatókat tartalmazó tényezőket elhagyjuk. Ez ekvivalens annak a feltételezésével, hogy a vizsgált változók korrelálatlanok (függetlenek).

Másodrendű közelítő formulához jutunk, ha a fenti összefüggésben csak a kettőnél több változóra vonatkozó korrelációs együttthatókat tartalmazó tagokat hagyjuk el. A közelítő formula a  $\pi_i(\underline{X})$ , ( $i=1,2$ ) valószínűségek becslt értékének meghatározásakor a változók közötti korrelációs kapcsolatokat is figyelembe veszi.

A másik három eljárás

- a Fisher-féle (2) lineáris (FL),
- a Marks-féle (3) ún. "legjobb" lineáris (LL), illetőleg
- a kvadratikus (QD)

eljárás binarizált változata.

A három eljárás abban az esetben optimális, ha a válaszvektorok populációkénti eloszlása többváltozós normális eloszlással közelíthető. Az általános használt Fisher-féle eljárás feltételezi, hogy az egyes populációkat jellemző kovariancia mátrixok egyenlőek. Ha a feltevés fennáll, akkor az LL, illetve a QD eljárás az FL eljárásra redukálódik.

Ha a feltevés nem áll fenn, akkor az LL, illetőleg a QD eljárások - a téves besorolások valószínűségét tekintve - általában kedvezőbb eredményeket szolgáltatnak. LL a hiba valószínűségét minimalizálja.

A vizsgált eljárások ( $\mu = B1, B2, FL, LL, QD$ ) az

$$L(\underline{X}, \mu) = \log \left[ \pi_2(\underline{X}, \mu) / \pi_1(\underline{X}, \mu) \right]$$

ún. log likelihood hányadost becsülik.

Az FL, LL, QD eljárások a log likelihood hányadost a válaszok lineáris, illetőleg kvadratikus függvényével közelítik. A B1, B2 eljárások esetében a hányados meghatározása a  $\pi_i(\underline{X}, \mu)$ , ( $i=1,2$ ,  $\mu=B1,B2$ ), multinomiális valószínűségek alapján történt. (5., 6. ábra)

### LINEÁRIS DISZKRIMINANCIA FÜGGVÉNY

$$L(\underline{X}, LDF) = \underline{X}' S' (\bar{\underline{X}}_2 - \bar{\underline{X}}_1) - \frac{1}{2} (\bar{\underline{X}}_2 - \bar{\underline{X}}_1)' S' (\bar{\underline{X}}_2 - \bar{\underline{X}}_1)$$

ahol:  $\underline{X}$  a vizsgált személy sz.kl.p.-ának vektora,

$\underline{X}'$  a sz.kl.p.-ok vektorának transzponáltja,

$\bar{\underline{X}}_1, \bar{\underline{X}}_2$  az egyes sz.kl.p.-ok populációnkénti várható értékének vektora,

$S'$  az összesített kovariancia mátrix inverze.

SZ.KL.P. = SZUBJEKTÍV KLINIKAI PANASZ + RR

5. ábra

### KVADRATIKUS DISZKRIMINANCIA FÜGGVÉNY

$$L(\underline{X}, QDF) = \frac{1}{2} \underline{X}' (S_1^{-1} - S_2^{-1}) \underline{X} + (\bar{\underline{X}}_2' S_2^{-1} \bar{\underline{X}}_1 - \bar{\underline{X}}_1' S_1^{-1} \bar{\underline{X}}_2) \underline{X} - \frac{1}{2} (\bar{\underline{X}}_2' S_2^{-1} \bar{\underline{X}}_2 - \bar{\underline{X}}_1' S_1^{-1} \bar{\underline{X}}_1)$$

ahol:  $\underline{X}$  a vizsgált személy sz.kl.p.-ának vektora,

$\underline{X}'$  a sz.kl.p.-ok vektorának transzponáltja,

$\bar{\underline{X}}_1, \bar{\underline{X}}_2$  az egyes sz.kl.p.-ok populációnkénti várható értékének vektora,

$S_1^{-1}, S_2^{-1}$  az egyes populációk kovariancia mátrixának inverze.

SZ.KL.P. = SZUBJEKTÍV KLINIKAI PANASZ + RR

6. ábra



Az egyének besorolására az ún. Bayes-féle szabályt használtuk:

Jelölje:  $B(\underline{X}, \mu)$ ,  $\mu = B1, B2, FL, LL, QD$   
annak a valószínűségét, hogy az  $\underline{X}$  válaszvektorral jellemzett személy a 2. (beteg) populációba tartozik, akkor

$$B(\underline{X}, \mu) = \begin{cases} 1 & \text{ha } L(\underline{X}, \mu) \\ 1 - \pi & \text{ha } L(\underline{X}, \mu) \\ 0 & \text{ha } L(\underline{X}, \mu) \end{cases} = \log \left[ \frac{\pi}{1 - \pi} \right]$$

ahol  $\pi$  annak az a priori valószínűsége, hogy a vizsgált személy az 1. (egészséges) populációba tartozik. A módszertani vizsgálat  $\pi = 0,5$  feltételezésével történt. (Ha az a priori valószínűségeket a vizsgált populáció sajátosságainak megfelelően állítjuk be, akkor a téves besorolások valószínűsége csökkenthető.)

Az egyes eljárások jellemző paramétereinek meghatározása, illetőleg az eljárások tesztelése egymástól független, véletlen minták alapján történt.

Az egyes kérdések diszkrimináló jellegének vizsgálatára az FL, illetve az LL eljárást használtuk. A két populáció Mahalanobis-féle általánosított távolsága:  $\Delta^2 = 2,2$ .

A 28 kérdésből 19-nél találtunk jó, illetőleg megfelelő elkülönítő képességet, 9 nem rendelkezett elfogadható mértékben ilyen tulajdonsággal (pl. fejfájás, bokaduzzanat, palpációérzés).

Jelentős adat, hogy a vérnyomásmérés diszkrimináló hatása 36 %. Az eredeti, teljes kérdéscsoport (28 adat) figyelembevételével történt számítógépes futtatás során (FL és QD) igen jó megbízhatósági értékeket kaptunk:

Függvény	Validitás	Szenzitivitás	Specificitás
FL	77 %	74 %	80 %
QD	80 %	75 %	87 %

A nem szignifikáns hatású kérdések elhagyása után a számítást az összes eljárás figyelembevételével megismételtük.

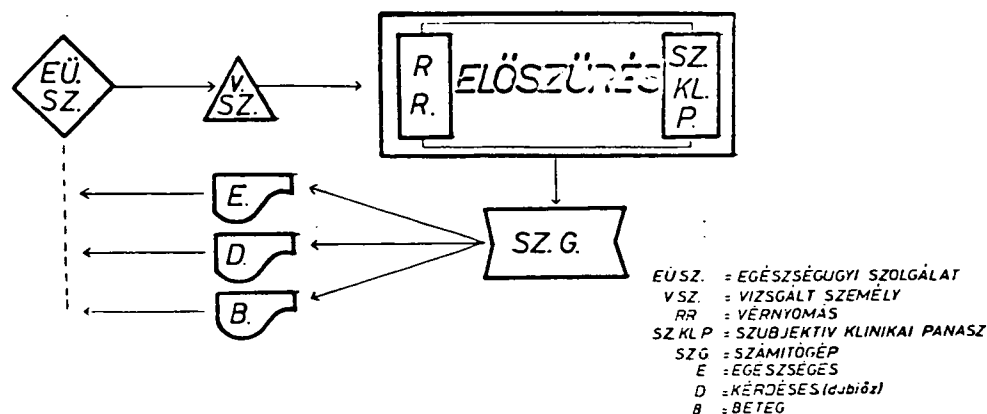
Az FL és a QD eljárások megbízhatósága lényegében nem változott. Az LL eljárás megbízhatósága valamivel jobb, mint az FL eljárásé, de a QD eljárás megbízhatóságát nem éri el.

A B1 eljárás megbízhatósága - a viszonylag laza korrelációs kapcsolatok ellenére - lényegesen kedvezőtlenebb, a B2 eljárásé valamivel kedvezőbb mint az FL eljárásé.

A vizsgált eljárásokra általában jellemző, hogy specifikitásuk lényegesen kedvezőbb, mint a szenzitivitásuk. A szűrővizsgálatokkal szemben támasztott alapvető követelményeknek viszont legjobban egy nagy szenzitivitású és megfelelő specifikitású eljárás felelne meg. Célkitűzéseinknek megfelelően az eredeti LL eljárást módosítottuk, vagyis rögzített specifikitás mellett, a szenzitivitást maximalizáltuk. 70 %-os specifikitás mellett az eredményül adódó diszkriminancia függvény validitása 76 %, szenzitivitása 80%, de a specifikitás rovására a szenzitivitás 90 %-ig növelhető. A váltoszám csökkenése lehetőséget nyújt arra, hogy a további vizsgálatok során iteratív becslési módszereket igénylő eljárások (maximum likelihood, ill. minimum  $X^2$  logit) alkalmazhatóságát is megvizsgáljuk.

Eddigi tapasztalataink és eredményeink alapján elmondhatjuk, hogy módszerünk alapján a cardiovascularis betegségcsoport számítógépes értékeléssel egybekapcsolt kérdőíves előszűrési rendszere a (7. ábrán) látható modell formájában kialakítható.

## SZÁMITÓGÉPES ÉRTÉKELESSLAL KAPCSOLT KÉRDŐÍVES ELŐSZÜRÉSI MODELLJE



7. ábra

Irodalom

- (1) Bahadur, R.R.: A Representation of the Joint Distribution of Responses to  $n$  Dichotomous Items, in H. Solomon, ed., Studies in Item Analysis and Prediction, Stanford: Stanford University Press, 1961, 158-68.
- (2) Fisher, R.A.: The Use of Measurements in Taxonomic Problems, Annals of Eugenics (September 1936) 179-88.
- (3) Marks, S.: Discriminant Functions When Covariance Matrices are Unequal, JASA, 69, 555-559.
- (4) Moore, D.H.: Evaluation of Five Discrimination Procedures for Binary Variables, JASA, 68, 399-404.
- (5) Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, New York: John Wiley and Sons, Inc., 1958.
- (6) Szilárd, I. és mtsai: Első tapasztalatunk a cardiovascularis betegségcsoport kérdőíves előszűrésével kapcsolatosan egy városi körzetben (előadás, Szeged, 1975. AMagyar Egészségügyi Szervezők Tudományos Egyesületének Kongresszusa).

